Estimation of Weekly Reference Evapotranspiration using Linear Regression and ANN Models

K.Chandra Sekhar Reddy[1], S.Aruna Jyothy[2], and P.Mallikarjuna[3]
[1] N.B.K.R.I.S.T., Department of Civil Engineering, Vidyanagar, Andhra Pradesh, India.
Email: vinod_harsha@yahoo.co.in
[2] Sri Venkateswara University, Department of Civil Engineering, Tirupati, India
Email: aruna_jyothy@yahoo.co.in
[3] Sri Venkateswara University, Department of Civil Engineering, Tirupati, India
Email : mallikarjuna_perugu@yahoo.co.in

Abstract— **The study investigates the applicability of linear regression and ANN models for estimating weekly reference evapotranspiration (ET_0) at Tirupati, Nellore, Rajahmundry, Anakapalli and Rajendranagar regions of Andhra Pradesh. The climatic parameters influencing ET_0 were identified through multiple and partial correlation analysis. The sunshine, temperature, wind velocity and relative humidity mostly influenced the study area in the weekly ET_0 estimation. Linear regression models in terms of the climatic parameters influencing the regions and, optimal neural network architectures considering these climatic parameters as inputs were developed. The models' performance was evaluated with respect to ET_0 estimated by FAO-56 Penman-Monteith method. The linear regression models showed a satisfactory performance in the weekly ET_0 estimation in the regions selected for the present study. The ANN (4,4,1) models, however, consistently showed a slightly improved performance over linear regression models.**

Index Terms—**Reference evapotranspiration, multiple linear regression, artificial neural network, performance evaluation**

I. INTRODUCTION

An accurate estimation of reference crop evapotranspiration (ET_0) is of paramount importance for designing irrigation systems and managing natural water resources. Numerous ET_0 equations have been developed and used according to the availability of historical and current weather data. These equations range in sophistication from empirical to complex equations. The FAO-56 Penman-Monteith (PM) equation [1] is widely used in recent times for ET_0 estimation. However, the difficulty in using this equation, in general, is the lack of accurate and complete data. In addition, the parameters in the equation potentially introduce certain amount of measurement and/or computational errors, resulting in cumulative errors in ET_0 estimates. Under these conditions, a simple empirical equation that requires as few parameters as possible and, results comparable with Penman-Monteith method is preferable. Owing to the difficulties associated with model structure identification and parameter estimation of the nonlinear complex evapotranspiration process, most of the models that have been developed may not yield satisfactory results. ANNs are capable of modelling complex nonlinear processes effectively extracting the relation between the inputs and outputs of a process without the physics being explicitly provided to them and also, they identify the underlying rule even if the data is noisy and contaminated with errors [2] and [3]. "Reference [4]" investigated the utility of ANNs for the estimation of daily ET_0 and compared the performance of ANNs with PM method. It was concluded that ANNs can predict ET_0 better than the conventional methods. "Reference [5]" examined the potential of artificial neural networks in estimating the actual evapotranspiration from limited climatic data and suggested that the crop evapotranspiration could be computed from air temperature using the ANN approach. "Reference [6]" showed that ANNs can be used for forecasting ET_0 with high reliability. "Reference [7]" derived solar radiation and net radiation based ET_0 equations using multi- linear regression technique and concluded that the equations performed better than the simplified temperature and/or radiation based methods for humid climates. "Reference [8]" tested the ANNs for estimating ET_0 as a function of maximum and minimum air temperatures and concluded that when taking into account just the maximum and minimum air temperatures, it is possible to estimate ET_0. "Reference [9]" tested the ANNs, to estimate ET_0 as a function of the maximum and minimum air temperatures in semiarid climate. While comparing with PM method, it was concluded that ANN methods are better for ET_0 estimates than the conventional methods. "Reference [10]" evaluated ANN models for daily ET_0 estimation under situations of presence of only temperature and relative humidity data. ANNs showed an improved performance over traditional ET_0 equations. "Reference [11]" developed generalized artificial neural network (GANN) based reference crop evapotranspiration models corresponding to FAO-56 PM, FAO-24 Radiation, Turc and FAO-24 Blaney-Criddle methods using the data from California Irrigation Management and Information System stations. It was concluded that the GANN models can be used directly to predict ET_0 under the arid conditions since they performed better than the conventional ET_0 estimation methods. "Reference [12]" compared weekly evapotranspiration ANN based forecasts with regard to a model based on weekly averages and found an improved performance of one week in advance weekly ET_0 predictions compared to the model based on means (mean year model).

© 2011 ACEE
DOI: 01.IJCEE.01.01.517

II. MATERIALS AND METHODS

The climatic data at Tirupati, Nellore, Rajahmundry, Anakapalli and Rajendranagar meteorological centers collected from the India Meteorological Department (IMD), Pune, India were used in the data analysis and model development. A part of the data was used for the purpose of development of models and the rest for validating the models developed. The resemblance of the statistical structure in terms of mean, variance and skewness of the calibration and validation data sets was ensured while making the division of the data into training and testing data sets. A brief description of the meteorological centers along with the data period is shown in TABLE I. In the present study, an attempt is made to develop simple linear and optimal neural network models considering the climatic parameters influencing the regions selected for the study for weekly ET_0 estimation. The study also compares the performance of proposed linear regression and ANN models.

III. MODEL DEVELOPMENT

The weekly reference evapotranspiration model at a meteorological center is developed using the climatic data at the center. The steps in the modelling include i) identification of meteorological parameters influencing the region, ii) development of the model and iii) performance evaluation of the model developed. The identification of meteorological parameters influencing the region is based on multiple and partial correlation analysis. The linear regression and ANN models are developed for the present study. The performance of the models is verified through selected performance evaluation criteria.

A. Linear Regression (LR) model

The objective of the model is the transfer of information among several variables observed simultaneously and the estimation of the dependent variable from the several other observed independent variables.

The weekly reference evapotranspiration (ET_0) at a meteorological center is expressed as a simple linear model as

$$ET_0 = C + a_1 X_1 + a_2 X_2 + \ldots \qquad (1)$$

where $a_1, a_2, \ldots$ and C are empirical constants and $X_1, X_2, \ldots$ are the meteorological parameters influencing the region. The multiple correlation analysis was carried out using STASTICA package.

B. Artificial Neural Network model (ANN)

A standard multilayer feed-forward ANN with logistic sigmoid function was adopted for the present study. A constant value of 0.1 for learning rate and a constant value of 0.9 for momentum factor were considered. The data were normalized in the range of (0.1, 0.9) to avoid any saturation effect. Error back propagation which is an iterative nonlinear optimization approach based on the gradient descent search method [13] was used during calibration. The calibration set was used to minimize the error and validation set was used to ensure proper training of the neural network employed such that it does not get overtrained. The performance of the model was checked for its improvement on each iteration to avoid overlearning. The optimal network corresponding to minimum mean squared error was obtained through trail and error process. Care was taken to avoid too few and too many neurons which can respectively cause difficulties in mapping each input and output in the training set and increase training time unnecessarily, in the process of determination of optimal number of hidden layers and nodes in each hidden layer. The process was carried out using MATLAB routines.

IV. PERFORMANCE EVALUATION CRITERIA

The performance evaluation criteria used in the present study are the coefficient of determination (R^2), root mean square error (RMSE) and efficiency coefficient (EC).

TABLE I.
BRIEF DESCRIPTION OF METEOROLOGICAL CENTERS

Meteorological center	Longitude (°E)	Latitude (°N)	Altitude (m)	Mean daily relative humidity (%)	Mean daily temperature (°C)	Mean daily wind velocity (kmph)	Mean daily sunshine hours (hr)	Mean daily vapour pressure (mm of Hg)	Mean annual rainfall (mm)	Data period	
										Trianing	Testing
Tirupati	79° 05'	13° 05'	161.0	59.5	28.2	7.9	6.8	17.6	1100	1992-98	199-01
Nellore	79° 59'	14° 22'	19.0	77.3	25.6	6.3	7.3	20.3	1170	1983-97	1998-03
Rajahmundry	81° 46'	17° 00'	14.0	70.9	27.8	6.3	7.1	20.4	1160	1990-97	1998-01
Anakapalli	83° 01'	17° 38'	25.0	71.9	27.9	4.6	7.1	20.6	1190	1980-94	1995-01
Rajendranagar	78° 23'	17° 19'	536.0	61.8	26.2	7.3	8.0	14.9	920	1978-88	1989-93

A. Coefficient of Determination (R^2)

It is the square of the correlation coefficient (R) and the correlation coefficient is expressed as

$$R = \frac{\sum_{i=1}^{n} (y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\left[\sum_{i=1}^{n}(y_i - \bar{y})^2 \sum_{i=1}^{n}(\hat{y}_i - \bar{\hat{y}})^2\right]^{1/2}} \times 100 \qquad (2)$$

where y_i and $\hat{y}_i$ are the observed and estimated values respectively and, $\bar{y}$ and $\bar{\hat{y}}_i$ are the means of observed and estimated values and n is the number of observations. It measures the degree of association between the observed and estimated values and indicates the relative assessment of the model performance in dimensionless measure.

© 2011 ACEE
DOI: 01.IJCEE.01.01.517

B. Root Mean Square Error (RMSE)

It yields the residual error in terms of the mean square error and is expressed as [14].

$$\text{RMSE}=\sqrt{\frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{n}} \qquad (3)$$

C. Efficiency Coefficient (EC)

It is used to assess the performance of different models [15]. It is a better choice than RMSE statistic when the calibration and verification periods have different lengths [16]. It measures directly the ability of the model to reproduce the observed values and is expressed as

$$\text{EC} = \left(1-\frac{F}{F_0}\right)\times 100 \qquad (4)$$

where $F_0 = \sum_{i=1}^{n}(y_i-\bar{y})^2$ and $F = \sum_{i=1}^{n}(y_i-\hat{y}_i)^2$

A value of EC of 90% generally indicates a very satisfactory model performance while a value in the range 80-90%, a fairly good model. Values of EC in the range 60-80% would indicate an unsatisfactory model fit.

V. RESULTS AND DISCUSSION

The multiple correlation analysis has been carried out to identify the climatic parameters influencing weekly ET_0 in the regions selected for the study. It may be observed from multiple and partial correlation coefficients presented in TABLE II that the sunshine hours, temperature, wind velocity and relative humidity mostly influence the regions in the weekly ET_0 estimation. Linear weekly ET_0 regression models at the centers have been developed as presented in TABLE III. The weekly ET_0 estimation at the meteorological centers has also been carried out using different artificial neural network architectures with input nodes ranging from one to four and, varying the number of nodes in the hidden layer. The ANNs with four input nodes and a hidden layer with four nodes i.e. ANN (4,4,1) have been identified as optimal architectures. The performance indices of linear regression (LR) and ANN (4,4,1) models on comparison of the results with those of FAO-56 Penman-Montieth method are presented in TABLE IV. It may be observed from the results presented in TABLE IV that the values of R^2 and EC of LR models indicate a very satisfactory performance. However, the performance has improved marginally with optimal artificial neural network architectures.

TABLE II
MULTIPLE AND PARTIAL CORRELATION COEFFICIENTS

Meteorological center	Multiple correlation coefficient, Independent variable omitted							Partial correlation coefficient					
	----	T	S	W	RH	VP	RF	T	S	W	RH	VP	RF
Tirupati	0.9921	0.9843	0.9639	0.9491	0.9909	0.9919	0.9919	0.7034	0.8821	0.9173	0.3624	0.1568	0.1568
Nellore	0.9785	0.9736	0.8586	0.9412	0.9755	0.9783	0.9781	0.4285	0.9155	0.7920	0.3480	0.0955	0.1344
Rajahmundry	0.9841	0.9767	0.9513	0.9649	0.9821	0.9840	0.9840	0.5613	0.8173	0.7366	0.3329	0.0787	0.0787
Anakapalli	0.9791	0.7385	0.8600	0.9301	0.9768	0.9789	0.9790	0.9534	0.9171	0.8327	0.3132	0.0968	0.0686
Rajendranagar	0.9755	0.8816	0.9521	0.9161	0.9697	0.9753	0.9754	0.8847	0.6945	0.8360	0.4348	0.0894	0.0634

TABLE III
LINEAR REGRESSION MODELS

Meteorological Center	Regression Equation
Tirupati	$ET_0 = -0.939 + 0.199\,T + 0.210\,S + 0.181\,W - 0.038\,\mathbf{RH}$
Nellore	$ET_0 = -1.535 + 0.218\,T + 0.239\,S + 0.125\,W - 0.035\,\mathbf{RH}$
Rajahmundry	$ET_0 = -2.631 + 0.246\,T + 0.226\,S + 0.118\,W - 0.029\,\mathbf{RH}$
Anakapalli	$ET_0 = -3.907 + 0.239\,T + 0.236\,S + 0.148\,W - 0.012\,\mathbf{RH}$
Rajendranagar	$ET_0 = -3.219 + 0.248\,T + 0.240\,S + 0.140\,W - 0.023\,\mathbf{RH}$

© 2011 ACEE
DOI: 01.IJCEE.01.01.517

TABLE IV
PERFORMANCE INDICES OF LR AND ANN (4,4.1) MODELS

Meteorological center	Variables	Slope of the scatter plot		Intercept of the scatter plot		R^2		RMSE (mm)		EC (%)	
		LR	ANN (4,4,1)	LR	ANN (4,4,1)	LR	ANN (4,4,1)	LR	ANN (4,4,1)	LR	ANN (4,4,1)
Tirupati	RH,T,W,S	0.9650	1.0051	0.2105	0.2609	0.9863	0.9930	0.18	0.13	98.63	99.30
Nellore	RH,T,W,S	0.9785	0.8545	0.0552	0.3277	0.9672	0.9850	0.25	0.17	96.72	98.50
Rajahmundry	RH,T,W,S	0.9773	0.9012	0.1686	0.1657	0.9643	0.9671	0.22	0.21	96.43	96.71
Anakapalli	RH,T,W,S	0.9869	1.0484	0.0240	-0.1270	0.9549	0.9415	0.20	0.23	95.49	94.15
Rajendranagar	RH,T,W,S	0.9166	1.0016	0.3410	-0.0052	0.9127	0.9888	0.40	0.14	91.27	98.88

The values of RMSE of ANN(4,4,1) models have also reduced slightly. This may be due to the fact that weekly average ET_0 values do not exhibit much of nonlinearity. The scatter plots (not shown in the paper) of ET_0 values estimated using Penman-Montieth method against those estimated using LR and ANN (4,4,1) models respectively. The nearly unit slope and zero intercept of scatter plots (TABLE IV) indicate the closeness of ET_0 values with those of PM method. "Fig. 1" presents the comparison of performance of LR and ANN (4,4,1) models against PM method during testing period. The study reveals that the simple linear regression models proposed may be adopted satisfactorily in the weekly ET_0 estimation at the centers selected for the present study and, the accuracy in the ET_0 estimation may further be improved using ANN (4,4,1) models.

VI. CONCLUSIONS

The climatic parameters such as sunshine hours, temperature, wind velocity and relative humidity mostly influenced weekly ET_0 estimation at Tirupati, Nellore, Rajahmundry, Anakapalli and Rajendranagar regions of Andhra Pradesh. The linear regression models proposed in terms of the climatic parameters influencing the regions performed satisfactorily in the weekly ET_0 estimation. The optimal ANN models proposed showed a marginal improvement over linear regression models. The linear regression models may therefore be adopted for weekly ET_0 estimation in the regions with reasonable degree of accuracy and, the accuracy may slightly be improved with ANN architectures.

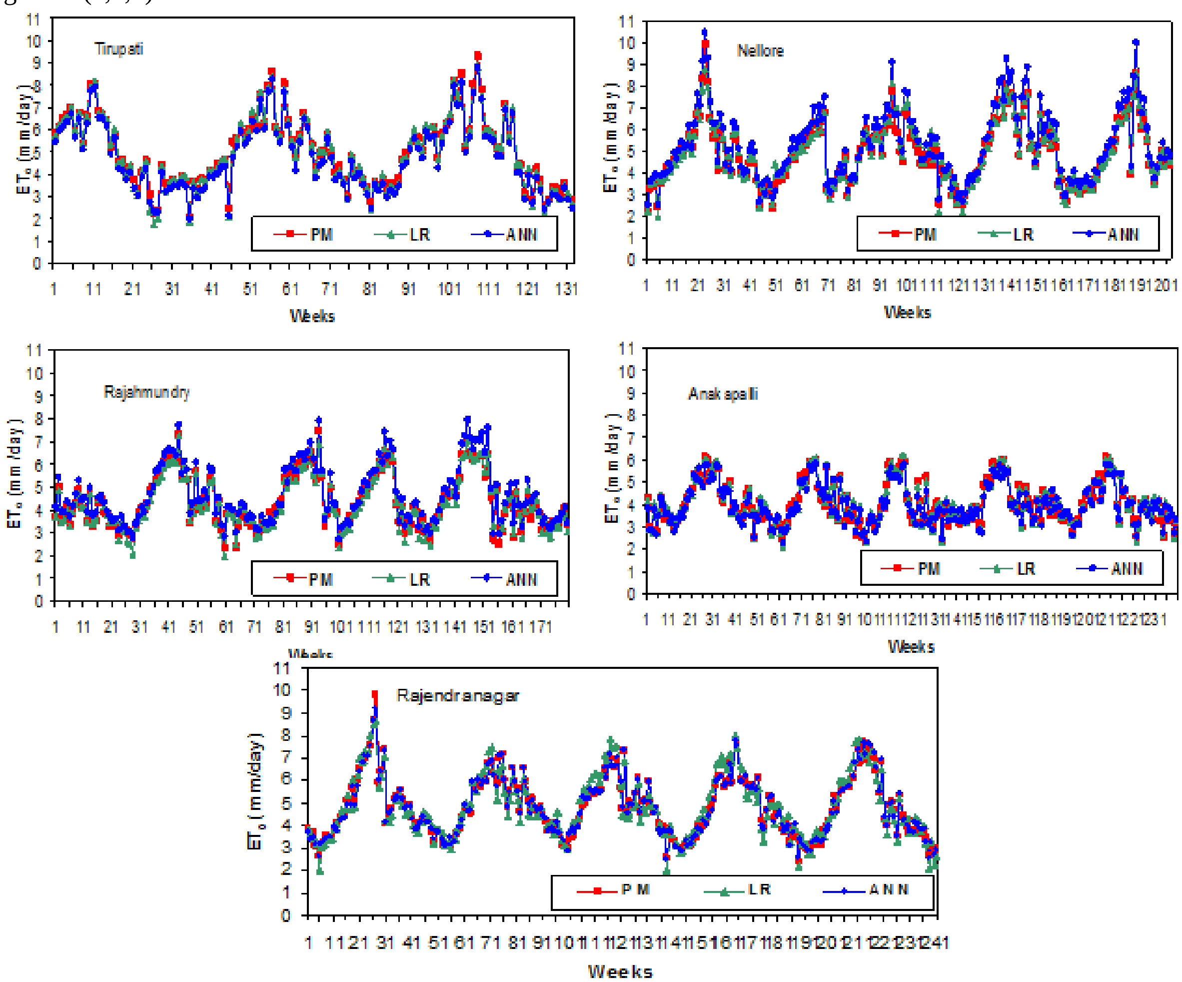

Figure 1. Comparison of average weekly ET_O values estimated using LR and ANN models with those estimated by Penman-Monteith method during testing period

© 2011 ACEE
DOI: 01.IJCEE.01.01.517

REFERENCES

[1] R. G. Allen, L. S Pereira, D. Raes, and M.Smith, "Crop Evapotranspiration Guidelines for Computing Crop Water Requirements." *FAO Irrigation and Drainage Paper 56*, FAO, Rome, 1998.

[2] ASCE Task Committee on Application of Artificial Neural Networks in Hydrology. "Artificial Neural Networks in Hydrology. I: Preliminary concepts." *J. Hydrol. Eng.*, 5(2), pp. 115-123, 2000a.

[3] ASCE Task Committee on Application of Artificial Neural Networks in Hydrology."Artificial Neural Networks in Hydrology. II: Hydrologic applications." *J. Hydrol. Eng.*, 5(2), pp. 124-137, 2000b.

[4] M. Kumar, N. S. Raghuwanshi, R. Singh, W. W.Wallender, and W. O. Pruitt. "Estimating Evapotranspiration using Artificial Neural Network." *J. Irrig. Drain. Eng.*, 128(4), pp. 224-233, 2002.

[5] K. P. Sudheer, A. K. Gosain, and K. S. Ramasastri, "Estimating Actual Evapotranspiration from Limited Climatic Data using Neural Computing Technique." *J. Irrig. Drain. Eng.*, 129(3), pp. 214-218, 2003.

[6] S. Trajkovic, B.Todorovic, and M. Stankovic, **"**Forecasting of Reference Evapotranspiration by Artificial Neural Networks." *J. Irrig. Drain. Eng.*, 129(6), pp.454-457, 2003.

[7] S. Irmak, A. Irmak, R. G. Allen, and J. W. Jones, "Solar and Net Radiation-Based Equations to Estimate Reference Evapotranspiration in Humid Climates." *J. Irrig. Drain. Eng.* 129(5), pp.336-347, 2003.

[8] S. S.Zanetti, E. F. Sousa, V. P. S.Oliveira, F. T.Almeida, and S. Bernardo,"Estimating Evapotranspiration using Artificial Neural Network and Minimum Climatological Data." *J. Irrig. Drain. Eng.*, 133(2), 83-89, 2007.

[9] A.R. Khoob, "Artificial Neural Network Estimation of Reference Evapotranspiration from Pan Evaporation in a Semi-Arid Environment." *J. Irrig. Sci.*, 27(1), pp.35-39, 2008.

[10] G. Landeras, A. O. Barredo, and J. J. Lopez, "Comparison of Artificial Neural Network Models and Empirical and Semi-Empirical Equations for Daily Reference Evapotranspiration Estimation in the Basque Country (Northern Spain)."*Agricultural Water Management*, 95(5), pp. 553-565, 2008.

[11] M. Kumar, N. S. Raghuwanshi and R. Singh, "Development and Validation of GANN model for Evapotranspiration Estimation." *J. Hydrol. Eng.*, 14(2), pp.131- 140, 2009.

[12] G. Landeras, A. O. Barredo, and J. J. Lopez. "Forecasting Weekly Evapotranspiration with ARIMA and Artificial Neural Network Models." *J. Irrig. Drain. Eng.*, 135(3), pp.323-334, 2009..

[13] D. E. Rumelhart, G. E.Hinton, and R. J. Williams. "Learning Internal Representations by Error Propagation." Parallel Distributed Processing, *MIT press*, *Cambridge*, *MA*, 1, pp. 318-362, 1986.

[14] P. S Yu., C. L Liu and T. Y. Lee, "Application of Transfer Function Model to a Storage Runoff Process." *Stochastic and Statistical Methods in Hydrology and Environmental Engineering*, 3, pp. 87 – 97, 1994.

[15] J. E. Nash, and J. V. Sutcliffe, "River Flow Forecasting Through Conceptual Models- I: A Discussion of Principles." *J. Hydrol.*, 10(3), pp. 282-290, 1970.

[16] G. C. Liang, O. Connor, and R. K. Kachroo, "A Multiple-Input Single-Output Varia R. K. Ble gain Model." *J. Hydrol.*, 155(1-2), pp. 185-198, 1994.

© 2011 ACEE
DOI: 01.IJCEE.01.01.517